iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** III **Month of publication:** March 2024

DOI: https://doi.org/10.22214/ijraset.2024.59610

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

Using Machine Learning for Social Network Spammer Detection & Fake User Identification

A. Deepika[1], P. Sruthi[2], K. Srikar[3], S. Sai Ganesh[4], P. Kranthi Kumar[5]
[1]*Assistant Professor, Department of CSE (Artificial Intelligence & Machine Learning), CMR College Of Engineering & Technology, Hyderabad, Telangana, India*
[2] *Associate Professor, Department of CSE (Artificial Intelligence & Machine Learning), CMR College Of Engineering & Technology, Hyderabad, Telangana, India*
[3, 4, 5]*Student, Department of CSE (Artificial Intelligence & Machine Learning), CMR College Of Engineering & Technology, Hyderabad, Telangana, India*

Abstract: ***In this research, various methods for Twitter spam detection are identified, and these methods are categorized into multiple categories to offer a taxonomy. For categorization, we have found four ways to report spammers which can be useful in spotting phony user IDs. The following criteria can be used to identify spammers:***

1) ***Phony content;***
2) ***URL-based spam detection;***
3) ***Spam detection in popular subjects;***
4) ***Fake identification of users.***

Keywords: Spam Detection, Extreme Learning Machines, Tweets.

I. INTRODUCTION

Using the Internet to find any type of data from any source worldwide has become very commonplace. Users are able to gather a vast amount of data and information about other users because to the rising demand for social media sites. Large amounts of data that are accessible on these websites also attract fraudulent users. Twitter has quickly developed into a real-time online information source regarding individuals. Users can share virtually anything on Twitter, an Online Social Network (OSN), including news, opinions, as well as their moods. Arguments about a variety of subjects, including politics, current events, and current affairs, can arise. A user's followers receive anything they tweet immediately, enabling them to disseminate the information they receive to a far larger audience. Studying and analysing user behaviour on online social networks has become more important as OSNs have developed. The scammers have the ability to effortlessly deceive a large number of individuals who lack knowledge about OSNs. Additionally, there's a need to stop and regulate those who use OSNs solely for the purpose of spamming other people's accounts with adverts. Researchers have recently become interested in the identification of spam on social networking platforms. Keeping social networks secure requires a challenging task: spam detection. To protect users from various dangerous assaults and to maintain their security and privacy, it is critical to identify spam on OSN websites. In the actual world, the community is severely destroyed by these risky tactics used by spammers. Twitter spammers aim to disseminate false information, rumours, fake news, and impromptu messages, among other things. Spammers use adverts and other methods to fund several mailing lists in order to accomplish their destructive goals. They then send out spam messages at random to promote their interests. The original users—known as non-spammers—are disrupted by these activities. Additionally, it damages the OSN platforms' reputation. In order to counteract spammers' destructive activities, it is imperative that a system for identifying spammers be developed. Numerous studies in the field of identifying Twitter spam have been conducted. A few surveys were additionally conducted on phony identities of users from Twitter in order to cover the state-of-the-art at this time. A study of novel approaches and strategies for Twitter spam detection is offered by Tinman et al. The survey mentioned above offers a comparison of the available methods. However, the authors also carried out a survey regarding the various actions that spammers on the social network Twitter take. Additionally, the study offers a survey of the literature acknowledging the presence of spammers on the social network Twitter. There remains a void in the body of literature despite all the studies that have been done. Therefore, we analyses the state-of-the-art in Twitter false user identification and spammer detection to close the gap. This poll also aims to provide a thorough overview of current advancements in the field and provides a taxonomy of Twitter spam detection techniques. This research aims to discover several methods of Twitter spam detection and to offer a taxonomy by grouping these methods into various categories.

For categorization, we have found four ways to report spammers which can be useful in spotting phony user IDs. The following criteria can be used to identify spammers: (i) phony content; (ii) URL-based spam detection; (iii) spam detection in popular subjects; and (iv) fake identification of users. offers a comparison of current methods and aids users in appreciating the importance and efficacy of the suggested approaches while also comparing their objectives and outcomes.

II. RELATED WORK

A. *''Twitter fake Account Detection,''*

Social networking sites like Facebook and Twitter are used by millions of individuals every day, and the interactions they have there have a big impact on their life. The widespread use of social networking has given rise to several worries, one of which is the potential for users to be duped by false accounts, which can cause harmful content to proliferate. In the actual world, this circumstance has the potential to seriously harm society. We report a classification technique in our work for identifying phony Twitter accounts. After utilizing a supervised discretization method called Entropy Minimization Discretization (EMD) on numerical features, we reprocessed our dataset and examined the outcomes of the Naïve Bayes algorithm.

B. *''Detecting Spammers on Twitter,''*

Internet users increasingly frequently utilize social networks to read news, discuss current events, and stay in touch with friends and family. People are spending more time on well-known social media platforms like Facebook, Twitter, and others, sharing and storing personal data.Malevolent users find this knowledge and the chance to connect with thousands of people intriguing. In order to accomplish their nefarious goals, they take advantage of the implicit trust relationships between users. For instance, they may send unsolicited communications to people who are not authorized, propagate fake news, or embed dangerous URLs into posts or tweets. Our aim in this study is to enhance the current spam detection techniques by examining the characteristics of Twitter spam users. We utilize a number of novel features (e.g., number of followers/followers, etc.) that are more reliable and efficient than already employed features for the purpose of identifying Twitter spammers. We assessed the suggested feature set through the use of well-known machine learning classification algorithms, including Decision Tree (DT), Naive Bayesian (NB), Random Forest (RF), k-Nearest Neighbours (k-NN), Logistic Regression (LR), Support Vector Machine (SVM), and extremes Gradient Boosting (XG-Boost). These classifiers' performances are assessed and contrasted using several assessment metrics. We evaluated the performance of our suggested method against the last four state-of-the-art methods. The experimental findings demonstrate that compared to current state-of-the-art methods, the suggested set of characteristics performs better.

C. *''An integrated approach for malicious tweets detection using NLP,''*

The detection of fraudulent user accounts has been the subject of numerous prior studies. Recently, social network research has focused on the detection of spammers or spams on Twitter. Nonetheless, we introduce a novel approach that focuses on two areas: first, identifying spam tweets without prior knowledge about the user; second, employing linguistic analysis to identify spam on Twitter related to popular subjects. Discussion subjects that are hot at that particular moment are known as trending themes. Thus, spammers profit from this expanding microblogging trend. Using linguistic tools, our work attempts to identify spam tweets. Before anything else, we gathered tweets about a variety of popular subjects and classified them as safe or dangerous. After labelling, we used language as a tool to extract several attributes based on language models. In addition, we assess the performance and categorize tweets as either spam or not. Therefore, by concentrating mostly on tweet analysis rather than user account analysis, our approach can be used to detect spam on Twitter.

D. *''Twitter spam detection: Survey of new approaches and comparative study"*

Twitter spam has always been a serious issue that is challenging to solve. Several defines and detection strategies have been put out by researchers thus far to shield Twitter users from spamming. When compared to the methods that were offered three years ago, a great deal of progress has been made in the last three years in terms of both detection efficiency and accuracy. Thus, we are inspired to develop a fresh survey concerning methods for detecting spam on Twitter. Three sections make up this survey: 1) A review of the literature on the state-of-the-art: this section offers in-depth analysis (such as feature selection biases and taxonomies) and debate (such as advantages and disadvantages of each common approach); 2) Comparative studies: to gain a quantitative grasp of current procedures, we will assess how well different common methods perform on a universal testbed (i.e., the same datasets and ground facts); 3) Open issues: The concluding section enumerates the outstanding problems with the state-of-the-art Twitter spam detection methods.

Both academics and industry place a high value on finding solutions to these open problems. Those who are looking for a thorough understanding of this topic to build new approaches, as well as those who do not have competence in this area, may read this survey.

III. METHODOLOGY

Description of 4 techniques to detect tweet is spam or normal.

The presented techniques are also compared based on various features, such as user features (retweets, tweets, followers etc.), content features (tweet content messages).

1) *Fake Content:* If the number of followers is low in comparison with the number of followings, the credibility of an account is low and the possibility that the account is spam is relatively high. Likewise, feature based on content includes tweets reputation, HTTP links, mentions and replies, and trending topics. For the time feature, if many tweets are sent by a user account in a certain time interval, then it is a spam account.
2) *Spam URL Detection:* The user-based features are identified through various objects such as account age and number of user favourites, lists, and tweets. The identified user-based features are parsed from the JSON structure. The quantity of (i) retweets, (ii) hashtags, (iii) user mentions, and (iv) URLs, on the other hand, are among the tweet-based features. We will use the Naïve Bayes machine learning technique to determine whether or not tweets contain spam URLs.
3) *Detecting Spam in Trending Topic:* In this technique tweets content will be classified using Naïve Bayes algorithm to check whether tweet contains spam or non-spam words. This algorithm will check for spam URL, adult content words and duplicate tweets. If Naïve Bayes detect tweet as SPAM, then it will return 1 and if not detected any SPAM content, then Naïve Bayes will return 0.
4) *Fake User Identification:* These attributes include the number of followers and following, account age etc. Alternatively, content features are linked to the tweets that are posted by users as spam bots that post a huge number of duplicate contents as contrast to non-spammers who do not post duplicate tweets. In this technique features (following, followers, tweet contents to detect spam or non-spam content using Naïve Bayes Algorithm) will be extracted from tweets and then classify those features with Naïve Bayes Algorithm as spam or non-spam. Later this feature will be train with random forest algorithm to determine account is fake or non-fake. All extracted features will be saved inside features.txt file. Naïve Bayes classifier saved inside 'model' folder.

IV. FLOW CHART

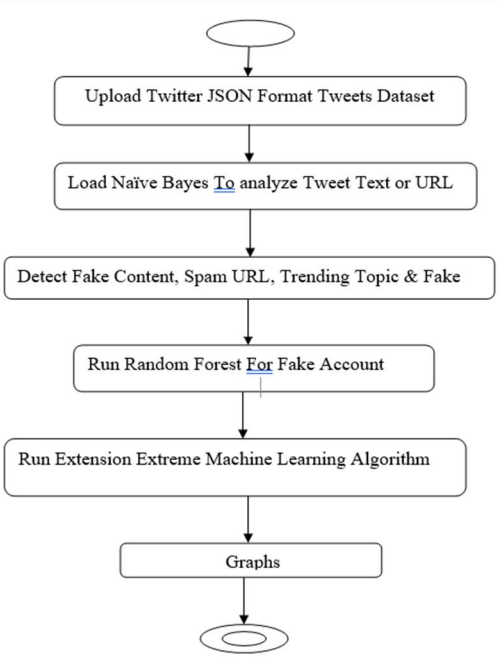


Fig-1: Flow Chart

V. RESULT AND DISCUSSION

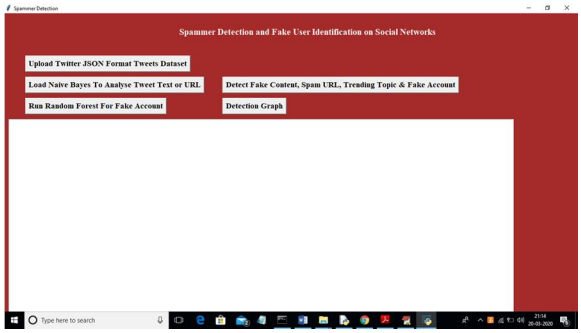


Fig-1: GUI

In above screen click on 'Upload Twitter JSON Format Tweets Dataset' button and upload tweets folder

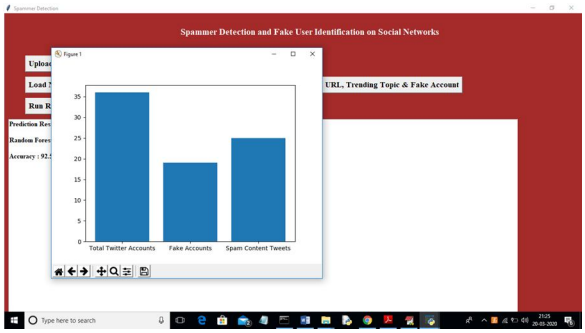


Fig-2: Account Analysis

In above graph x-axis represents total tweets, fake account and spam words content tweets and y-axis represents count of them.

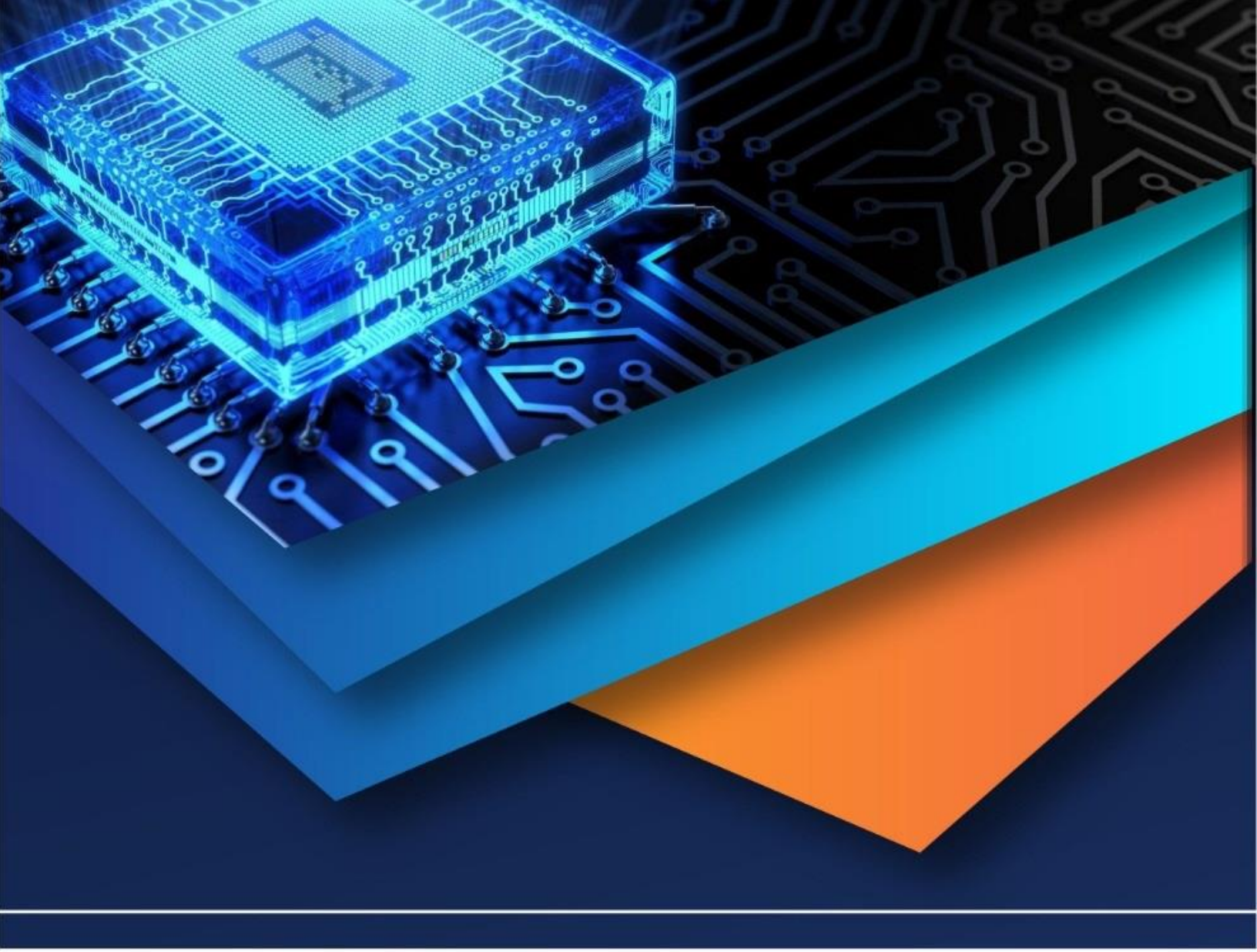
VI. CONCLUSION

In this work, we conducted an evaluation of methods for identifying Twitter spammers. Furthermore, a taxonomy of spam detection strategies was showcased, which was divided into four categories: spam identification in trending topics, false content detection, URL-based spam detection, or fake user detection methods. A number of variables, including user, content, graph, structure, and temporal aspects, were also used to compare the approaches that were presented. Additionally, the methods were contrasted with regard to the datasets and objectives they employed. Researchers should be able to get up-to-date information on Twitter spam detection methods in one convenient location with the aid of the review that has been presented. There are still certain unanswered questions that need to be given a lot of research effort, even with the development of successful and efficient methods for Twitter spam detection or fake user identification [34]. Below is a quick summary of the issues: Because of the grave consequences that false information can have on both an individual and a societal level, it is important to investigate the topic of false news identification in social media networks [25]. Finding rumour sources in social media is another related topic which is worth looking into. More advanced techniques, such as social network-based approaches, can be used because of their demonstrated efficacy, even if a few studies using statistical methods are currently being carried out to identify the sources of rumours.

REFERENCES

[1] B. Rechain, Ö. Aktaş, D. Kilinç, and C. Akyol, ''Twitter fake account detection,'' in Proc. Int. Conf. Compute. Sci. Eng. (UBMK), Oct. 2017, pp. 388–392.
[2] F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida, ''Detecting spammers on Twitter,'' in Proc. Collaboration, Electron. Messaging, AntiAbuse Spam Conf. (CEAS), vol. 6, Jul. 2010, p. 12.
[3] S. Ghadge, and M. Chavan, ''An integrated approach for malicious tweets detection using NLP,'' in Proc. Int. Conf. Inventive Common. Compute. Technol. (ICICCT), Mar. 2017, pp. 435–438.
[4] T. Wu, S. Wen, Y. Xiang, and W. Zhou, ''Twitter spam detection: Survey of new approaches and comparative study,'' Compute. Secure., vol. 76, pp. 265–284, Jul. 2018.
[5] S. J. Soman, ''A survey on behaviours exhibited by spammers in popular social media networks,'' in Proc. Int. Conf. Circuit, Power Compute. Technol. (ICCPCT), Mar. 2016, pp. 1–6.

[6] A. Gupta, H. Lamba, and P. Kumaraguru, ‘‘1.00 per RT #BostonMarathon # precombustion: Analysing fake content on Twitter,’’ in Proc. crime Researchers Summit (ecrus), 2013, pp. 1–12.
[7] F. Concone, A. De Paola, G. Lo Re, and M. Morana, ‘‘Twitter analysis for real-time malware discovery,’’ in Proc. AEIT Int. Annu. Conf., Sep. 2017, pp. 1–6.
[8] N. Shariq, M. Jalali, and M. H. Moat tar, ‘‘Detecting spam tweets in Twitter using a data stream clustering algorithm,’’ in Proc. Int. Conger. Technol., Common. Know. (ICTCK), Nov. 2015, pp. 347–351.
[9] C. Chen, Y. Wang, J. Zhang, Y. Xiang, W. Zhou, and G. Min, ‘‘Statistical features-based real-time detection of drifted Twitter spam,’’ IEEE Trans. Inf. Forensics Security, vol. 12, no. 4, pp. 914–925, Apr. 2017.
[10] C. Buntain and J. Golbeck, ‘‘Automatically identifying fake news in popular Twitter threads,’’ in Proc. IEEE Int. Conf. Smart Cloud (Smart Cloud), Nov. 2017, pp. 208–215.
[11] C. Chen, J. Zhang, Y. Xie, Y. Xiang, W. Zhou, M. M. Hassan, A. Allawi, and M. Arabian, ‘‘A performance evaluation of machine learning-based streaming spam tweets detection,’’ IEEE Trans. Compute. Social Syst., vol. 2, no. 3, pp. 65–76, Sep. 2015.
[12] G. Stafford and L. L. Yu, ‘‘An evaluation of the effect of spam on Twitter trending topics,’’ in Proc. Int. Conf. Social Compute., Sep. 2013, pp. 373–378.
[13] M. Mateen, M. A. Iqbal, M. Aleem, and M. A. Islam, ‘‘A hybrid approach for spam detection for Twitter,’’ in Proc. 14th Int. Bhuban Conf. Appl. Sci. Technol. (IBCAST), Jan. 2017, pp. 466–471.
[14] A. Gupta and R. Kaushal, ‘‘Improving spam detection in online social networks,’’ in Proc. Int. Conf. Cong. Compute. Inf. Process. (CCIP), Mar. 2015, pp. 1–6.
[15] F. Thaliana and M. Bauguess, ‘‘A model-based approach for identifying spammers in social networks,’’ in Proc. IEEE Int. Conf. Data Sci. Adv. Anal. (DSAA), Oct. 2015, pp. 1–9.
[16] V. Chauhan, A. Planica, V. Middha, A. Gupta, U. Bana, B. R. Prasad, and S. Agarwal, ‘‘Anomalous behaviour detection in social networking,’’ in Proc. 8th Int. Conf. Compute., Common. Newt. Technol. (ICCCNT), Jul. 2017, pp. 1–5.
[17] S. Jeong, G. Noh, H. Oh, and C.-K. Kim, ‘‘Follow spam detection based on cascaded social information,’’ Inf. Sci., vol. 369, pp. 481–499, Nov. 2016.
[18] M. Washa, A. Qurush, and F. Seeds, ‘‘Leveraging time for spammers detection on Twitter,’’ in Proc. 8th Int. Conf. Manage. Digit. Ecosystem., Nov. 2016, pp. 109–116.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY